MethodsNews

Summer 2015

Newsletter from the ESRC National Centre for Research Methods

IN THIS ISSUE

Data linkage: Ethical and social concerns

Rosalind Edwards, Christina Hughes and Malcolm Williams
Co-editors, International Journal of Social Research Methodology

Social scientists may be able to connect different sets of administrative information held about people technically, but should there be any rules about accessing and using the data? Such a concern extends to those areas of social research where there is open access to the data that people globally post on social media about themselves.

The accessibility of personal digital information, and the possibilities for linking together anonymised or pseudo-anonymised records of two or more large administrative or survey data sets, provide social scientists with unprecedented opportunities to transform our understandings of a range of social issues. Innovative technologies create new opportunities for knowledge but they also invade privacy and provide new channels for discrimination. The field is fraught with a range of ethical and social concerns. To stimulate and inform the important debate on these issues among social scientists, research funders and policy-makers, the International Journal of Social Research Methodology and ESRC National Centre for Research Methods held an innovation forum event, hosted by the Warwick Business School at The Shard, London.

The day began with Mandy Chessell, Master Inventor and Distinguished Engineer at IBM, delivering the NCRM Annual Lecture exploring the process of ethical decision-making about data linking in the business sector. Peter Elias and Rob Procter (both at Warwick) followed with presentations on the strengths and drawbacks of administrative data linkage and sharing social media information. Ian Dunt, editor of politics.co.uk, provided a political perspective on the policy-making process and civil liberties, while a panel consisting of Neil Serougi (independent researcher), Emma Uprichard (Warwick), David Martin (Southampton), Peter Smith (Southampton) and Rob Procter, identified opportunities and caveats in discussion with the audience.

The main deliberations revolved around:

- The benefits of data linkage:
  We can know about who people are, what they do, how and why they do it, for the betterment of society. There are also potential cost savings as data from many sources can be processed automatically and linked cheaply.
- Purpose and power:
  There is a business interest in the use of public data, both in terms of providing IT solutions for managing linkage, and in using the data for commercial purposes. The public, however, may feel that it is acceptable to link up administrative data only for the primary purpose for which was collected, such as health and social care provision. In particular, there is a concern that commercialisation of data and using data for non-intended purposes breaches rights to privacy. What is technically and legally possible is not necessarily what is ethically and socially right.
- Ethical governance
  In the British context there is no legal requirement for consent to administrative data linkage (albeit European legislation may impose this) and the conventional, institutional and professional research ethics that social scientists work with may not be sustainable in an age of big data and data science. The old practices are founded on principles of anonymity and informed consent, both of which are disrupted where big data, the World Wide Web and the internet of things have built-in linkage and are resistant to anonymisation. New ethical principles and practices need to be developed.

Possible ways forward for ethical data linkage practices and governance could include a new socio-technical architecture for data management involving a personal data store, raising understanding and building trust through public debate, and setting up an independent 'watchdog', data ombudsman, 'data parliament' or citizen's panel to ensure transparency, openness, a form of consent and appropriate use.

To watch the presentations and listen to the audio podcasts go to www.ncrm.ac.uk → resources → podcasts or videos.

Obtaining consent to link administration data for children in surveys

Tarek Al Baghal and Jonathan Burton, University of Essex

As fewer people are responding to surveys than ever before, the linking of administrative records to survey responses is a potentially important tool to obtain new, high-quality data while making surveys more efficient.

While initial research on adults suggest factors possibly influencing the consent decision and potential bias in records obtained, little research has been done on consent for children.

Given the possible benefits, a growing number of survey practitioners see linking administrative data to survey responses as an opportunity to improve data quality (and quantity) while reducing costs and easing interviewer and respondent burden. However, informed consent is frequently required to link survey and administrative data. Previous studies have largely focused on the consent process in surveys for adults; however, parents may also be asked to link their children's administrative records, which may be of particular interest in longitudinal studies. Inclusion of children's records allow for studying changes and outcomes over time for familial units, including intergenerational change.[1] These linked data may lack generalizability and may be biased, however, if consent is low and/or those who consented are different to those that did not.

To understand what factors possibly influence parents' decisions to consent for themselves and their children, a framework was developed to direct the research based on theories of survey participation generally.[2] These ideas have been identified as possibly important in understanding the consent decision, and include factors coming from the respondent, the respondent's environment, the interviewer, and the design of the survey. However, additional considerations need to be made when children are the subject of the consent request. For example, parenting style is an additional respondent factor that may not otherwise be considered. Further, it may not just be the parent's (i.e. the respondent's) characteristics influencing decisions, but also those of the children.

To examine and test this framework, our NCRM-funded research utilized data from *Understanding Society*: The United Kingdom Longitudinal Household Study. The survey is the largest longitudinal household survey in the UK, and at the first wave (in 2009-2010), respondents were asked if they consented to have their health and education records linked to their survey responses. For administrative reasons, only those born after 1981 and schooled in the UK were asked for linkage to their education records, while everyone was asked to link their NHS records. Parents were also asked for consent to link these same records for any children under 16, with education records only being requested if the child was school-aged. The rules regarding which parent was asked for children led to the mother nearly always being asked (95% of the time) and as such, only mothers were examined in this analysis.

Nearly all mothers gave the same consent response for all of their children, suggesting that in this instance what are important are mother-level factors, rather than child-level factors. This does not mean that future research should necessarily discount child characteristics. Other populations or survey contexts may lead to child-level factors having more of an impact. Mothers also consented for health records at a lower rate than for education records for both themselves and their children, while mothers consented for their children at lower rates than for themselves. However, a non-trivial number of mothers consented only for their children and not themselves (2.8%).

Analyses of consent outcomes show that mothers from minority ethnic groups are generally less likely to consent for either their children or themselves. Mothers who were harder-to-contact (and possibly more resistant to taking the survey) are more likely to refuse all consent requests. These respondents may feel that accepting the survey request is the extent of their willingness to participate. Improving strategies for reducing survey resistance and increasing willingness to share further may be possible through interviewer strategies developed while interacting with the respondent.[2] Building a rapport with the respondent may also lead to higher consent rates, suggested by the finding in this data that that longer interviews led to higher consent rates.

There was little impact of interviewer demographics, overall experience, and interviewer's achieved response rate and experience within *Understanding Society*, with no significant effects identified for mother or child health records. Why one consent request is apparently not affected by these interviewer success measures is unclear. Further exploration of which, when, and why interviewer traits are important is needed, which can then be used in interviewer recruitment and training.

In addition to identifying the factors related to consent, this study examined characteristics of children based on whether they were consented for or not to identify potential biases. There are a number of important demographic differences across children, with records less likely to be obtained among ethnic minorities and those in Southeast England and London. Responses to the youth survey of 10-15 year olds in *Understanding Society*, however, suggest little differences across several behaviours and attitudes, except for internet usage. While the lack of differences may be somewhat encouraging to users of linked data, the best way to minimise bias is to increase consent rates. For example, the differences in consent across ethnicities raise other concerns when the linked data are correlated with these demographics as many health outcomes may be.

References

1 Lightfoot, D. & Dibben, C. (2013) Approaches to linking administrative records to studies and surveys - a review. Administrative Data Liaison Service, University of St. Andrews. Retrieved from: http://www.adls.ac.uk/wp-content/uploads/Approaches-to-linking-administrative-records-to-studies-and-surveys-a-review.pdf

2 Groves, R. M. & Couper, M.P. (1998). Nonresponse in Household Interview Surveys. New York: Wiley.

Crime sensing with social media

Matt Williams and Pete Burnap, Cardiff University

The majority of individuals aged under twenty in the Western world were 'born digital' and will not recall a time without access to the Internet.

Combined with the migration of the 'born analogue' generation onto the Internet, fuelled by the rise of social media, we have seen the exponential growth of online spaces for the mass sharing of opinions and sentiments. These online spaces represent a socio-technical assemblage that creates a new public sphere enabling digital citizenship through which aspects of civil society are played out. No study of contemporary society can ignore this dimension of social life. New forms of digital online social data, handled by computational methods, allow social scientists to gain meaningful insights into contemporary social processes at unprecedented scale and speed. How we marshal these new forms of data present key challenges for the social sciences. Our NCRM Methodological Innovation Project sought to contribute to the methodological and capacity step-change needed to marshal and extract value from these new forms of data.

In the project *Social media and prediction: crime sensing, data integration and statistical modeling* we explored data fusion techniques to build a series of statistical models using heterogeneous datasets to gain insights from social media data to 'sense' offline crime patterns in London. We collected a corpus of 180 million UK geo-coded tweets covering a 12-month period using the COSMOS platform[1]. To reduce the size of the dataset to a sub-corpus of tweets related to crime and disorder communications we first developed a coding frame using existing interviews with victims of crime from the ESRC UK Data Archive. Extracts of interview data were prepared for crowd-sourced human coding via the CrowdFlower platform and gold standard annotations were derived. The human verified annotations were used to derive a lexicon that allowed us to classify the whole Twitter dataset. The resulting crime and disorder tweet sub-corpus was then also subject to further human annotation for validation. This verified corpus was used to develop a social media 'crime sensing' algorithm to automatically identify mundane references to crime and disorder in social media communications, using terms and phrases that were statistically likely to appear in content classified by human annotators. The algorithm was supplemented with emotive and affective terms using the WordNet Affect online lexical resource to identify content that would be suggestive of fear, distress and anxiety.

We explored a range of statistical methods for combining social media data with administrative (recorded crime) and curated data (census). Our dependent variables were measures of police recorded crime collected over a 12-month period provided by the Metropolitan Police Service. Given the desire to incorporate the temporal and spatial variability of police recorded crime and Twitter data with the static variables from the census, we used Random-Effects models that combine time-variant with time-invariant regressors. This meant that we could explore correlations between independent variables including tweets that have high temporal granularity and variability (every second) and census variables that have very low temporal granularity (every decade) with the dependent variable police recorded crime. The models included information from tweets relating to mentions of crime and disorder (such as criminal damage) and emotive states (such as anxiety).

Preliminary results indicate that models that include social media data improve the amount of variance explained in police recorded crime patterns, compared to models that include conventional crime predictors alone. This modelling technique may prove effective in sensing crime patterns ahead of conventional means. This project is an extension of our previous ESRC funded work on modelling the classification of racial tension[2,3] and propagation of cyberhate in social media[4,5,6]. We are now working with the Metropolitan Police Service via an ESRC Impact Acceleration Award to embed our computational and statistical models into their operational processes.

As part of the drive to up-skill social scientists in the area of big data analytics we have conducted a series of advanced workshops at the Web Science Trust Summer Schools at Singapore National University and in Southampton. We have provided training to Wales DTC doctoral students and we are currently developing a master's degree in Social Data Science at Cardiff as part of the University's new Social Data Science Lab.

References

1 Burnap, P. et al. 2014. COSMOS: Towards an integrated and scalable service for analysing social media on demand. International Journal of Parallel, Emergent and Distributed Systems 30:2, 80-100

2 Williams, M. L. et al. 2013. Policing cyber-neighbourhoods: Tension monitoring and social media networks. Policing and Society 23:461-481. 3.

3 Burnap, P. et al. 2015. Detecting Tension in Online Communities with Computational Twitter Analysis. Technological Forecasting and Social Change 95.

4 Burnap, P. et al. 2014. Tweeting the terror: modelling the social media reaction to the Woolwich terrorist attack. Social Network Analysis and Mining 4:206: 1-14.

5 Burnap, P. & Williams, M. L. 2015. Cyber hate speech on Twitter: An application of machine classification and statistical modeling for policy and decision making. Policy & Internet 7, 223-42.

6 Williams, M. L., and Burnap, P. 2015. Cyberhate on social media in the aftermath of Woolwich: A case study in computational criminology and big data. British Journal of Criminology, 1-28.

Tackling quantitative methods pedagogy

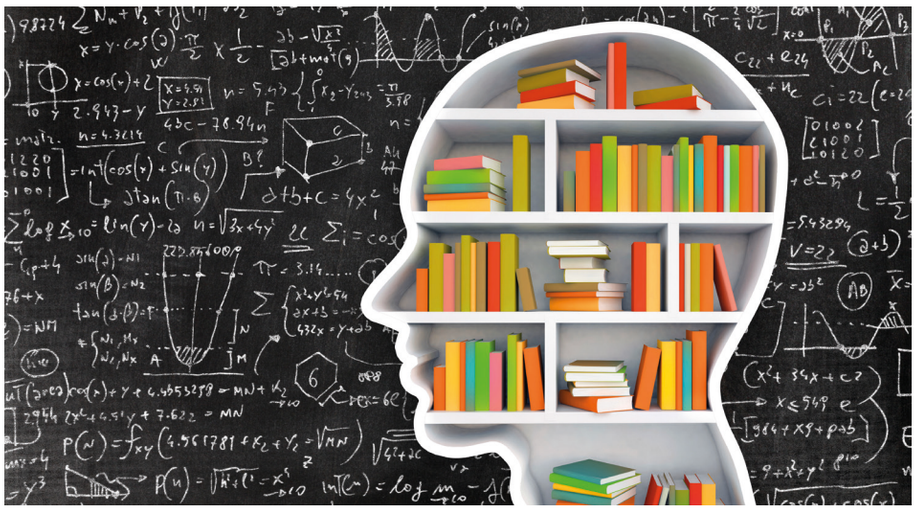
Kevin Ralston, John MacInnes, Vernon Gayle and Graham Grow, NCRM, University of Edinburgh

The challenge of training future cohorts of quantitative social science researchers to secure the 'pipeline' from school/ undergraduate study through postgraduate to postdoctoral research, depends not only on resources but on effective pedagogy[1].

A successfully functioning pipeline would generate a stream of social scientists capable of engaging with quantitative data and evidence. This could be important to maintain the relevance of the social sciences in a world where the ability to process, manage and analyse large volumes of data both opens up new opportunities to enhance knowledge, whilst, at the same time, presenting methodological challenges[2].

While there is research evidence on statistics teaching at school and university levels, most is USA-based and much of the literature is prescriptive, providing 'how to' recipes for classes[3]. There is little evidence-based work that addresses social science teaching specifically, where the challenge of poor maths skills and of a lack of confidence in applying them is often considered particularly acute. Indeed, although the current pedagogical literature is a useful resource, it is far from comprehensive and the pedagogical culture is limited[4].

In the context of many current initiatives (Nuffield Foundation Q-Step, evolving ESRC Doctoral Training Centres, Applied Quantitative Methods Network training,the British Academy Count Us In data skills strategy) NCRM has a unique opportunity to work with quantitative methods trainers and students to learn more about 'what works' in pedagogical and career development terms. Our research focus is on learning modes and achievements, motivation, student recruitment and retention. To enable this we are in the process of reviewing literature, evaluating tools and conducting secondary analysis as a prelude to primary data collection. This project provides one step towards developing a pedagogical culture based on evidenced and peer reviewed literature, which could provide a foundation on which a pipeline producing quantitative methods literate graduates and post-graduates could be constructed.



One possibly fruitful theme of the data collection for this project may focus upon student anxiety. The belief that social science undergraduates are apprehensive about their studies related to maths, statistics and quantitative methods in general is often cited in the literature[5,6]. By contrast, other research suggests the concept to be overstated, showing that 40% of sociology students in a single institution in the USA who responded to a survey report no angst[7] and only a slight majority report angst in a sample of students in England and Wales[8]. Tools such as the Statistical Anxiety Scale[9] and the Maths Anxiety Rating Scale[10,11] have been developed to measure the level of angst, with some comparisons between academic fields[12]. Our study could undertake a comprehensive comparison of anxiousness between social science disciplines and between social science and more numerate subjects, so that we might know whether social science students are significantly more 'frightened' of numbers than students in other disciplines or whether some social sciences are faring better than others in this respect.

References

1 Payne, G., Williams, M. (Eds.), 2011. Teaching Quantitative Methods: Getting the Basics Right. SAGE Publications Ltd, Los Angeles, Calif.

2 Savage, M., Burrows, R., 2007. The coming crisis of empirical sociology. Sociology 41, 885–899. doi:10.1177/0038038507080443

3 Gelman, A, Nolan, D., 2002. A probability model for golf putting. Teaching Statistics 24, 93-95.

4. Lewthwaite, S., Nind, M., 2015. Let's talk about pedagogy. MethodsNews Spring 2015. http://eprints.ncrm.ac.uk/3754/

5 Bridges, G.S., Pershing, J.L., Gillmore, G.M., Bates, K.A., 1998. Teaching quantitative research methods: A quasi-experimental analysis. Teaching Sociology 26, 14–28.

6 Paxton, P., 2006. Dollars and sense: Convincing students that they can learn and want to learn statistics. Teaching Sociology 34, 65–70.

7 DeCesare, M., 2007. "Statistics anxiety" among sociology majors: A first diagnosis and some treatment options. Teaching Sociology 35, 360–367.

8 Williams, M., Payne, G., Hodgkinson, L., Poade, D., 2008. Does British sociology count? Sociology students' attitudes toward quantitative methods. Sociology 42, 1003–1021.

9 Oliver, A., Sancho, P., Galiana, L., Cebria i Iranzo, M.A., 2014. New evidence on the Statistical Anxiety Scale (SAS). Anales de Psicologia 30, 151–157.

10 Bessant, K.C., 1995. Factors associated with types of mathematics anxiety in college students. Journal of Research in Mathematics Education 26, 327–345.

11 Suinn, R.M., Winston, E.H., 2003. The Mathematics Anxiety Rating Scale, a brief version: psychometric data. Psychological Reports 92, 167–173.

12 Hamza, E., Helal, A.M., 2013. Maths anxiety in college students across majors: A cross-cultural Study. Educationalfutures 5, 58-74.

The inquiry into the failure of the 2015 pre-election polls

Patrick Sturgis, NCRM, University of Southampton

The only people who woke up on the morning of May the 8th this year feeling worse than Ed Milliband and Nick Clegg were the pollsters.

Although it was clear during the campaign period that the exact vote share would be difficult to predict, the clear consensus from the polling data was that a hung parliament was a near certainty. In the end, of course, David Cameron returned to Downing Street with a 6.5% lead over Labour and a narrow but clear majority in the Commons, the first Conservative Prime Minister to achieve this since 1992. The 1992 election was also the last time that the polls got the result so spectacularly wrong, a pattern which some believe may prove to be more than just coincidence. The review into the 1992 polling miss concluded that the error was due to a combination of factors, notably 'late swing' and inaccurate population data for setting sample quotas. So what went wrong with the polls in 2015? At present it is too early to say but there are some likely contenders which the British Polling Council/Market Research Society Inquiry will be considering.

First, every pollster knows their predictions come with a 'margin of error' due to sampling variability, hence the result may be a few percentage points above or below the estimate from any one poll. Sampling variability can't be dismissed entirely but the size of the error and its conformity across polling organisations renders it very unlikely as the sole or even a notable contributory factor.

A second possibility is 'late swing', people changing their minds about which party they will vote for late in the campaign, after the final polls have taken place. This is certainly plausible and at least one prominent pollster has advanced this as the likely key explanation in 2015. However, polls taken on 7th May by a number of polling organisations showed no evidence of vote switching between the final poll and the election day. So, while late swing may have been part of the problem, it seems unlikely that it will account for much of it.

There is also the well-known 'shy Tory' effect, people apparently too embarrassed to admit they are going to vote Conservative and lying to pollsters about their intentions. This was a favoured explanation in 1992, anecdotally at least. On the face of it, it seems less relevant this time, as most of the polls these days are carried out on the internet, rather than face-to-face as they were in 1992. There is no reason to think that respondents should be embarrassed about clicking a mouse to indicate who they intend to vote for.

It may also be the case that pollsters drew their samples in ways that over-represented Labour voters at the expense of Tories and over-represented likely voters at the expense of those less likely to vote. There are good reasons to expect that these factors would have been evident in 2015 and the pollsters took a range of measures to try to correct for them. However, it is difficult to know *a priori* exactly which adjustments should be made and bias arising from sample composition and from adjustment measures are likely to have played at least some part in what went wrong.

A new possibility this time around relates to a phenomenon referred to as 'herding'. Herding is when poll estimates converge on a consensus estimate and, in 2015, the consensus estimate turned out to be wrong. There is some debate, much of it rather acrimonious, about how herding behavior arises. Although some pollsters may deliberately align their estimates with the majority position out of fear of being wrong, the more likely explanation is that the herding arises 'unconsciously' through the effect of prior beliefs about the likely result on adjustment decisions. Pollsters having to make important decisions about how to adjust their raw data have no way of being sure how to do this correctly in advance. It is feasible that these micro decisions may be influenced by prior beliefs about the likely outcome and these beliefs are themselves likely to derive from existing poll estimates. This pattern induces a circularity which causes estimates to converge.

These potential explanations, as well as others which may emerge, will be considered by the inquiry which is due to report in March 2016.

NCRM Director, Professor Patrick Sturgis chairs the BPC/MRS inquiry into the 2015 pre-election polls. Details of the Inquiry can be found at www.ncrm.ac.uk/polling.

Does audio-recording open-ended survey questions improve data quality?

Rebekah Luff and Patrick Sturgis, NCRM, University of Southampton

While the vast majority of survey questions use fixed response options, there is a long tradition in survey research of obtaining less structured data from respondents. These types of 'verbatim' responses require respondents to express their thoughts about a given topic 'in their own words'. Open-ended questions potentially provide richer data than closed-format questions and allow people to frame an issue in their own terms, rather than in those selected in advance by the researcher.

However, in face-to-face surveys, interviewers typically type open-ended responses as the respondent speaks, something which can be difficult to do fully and accurately. The quality of verbatim responses can therefore be quite variable, depending to a large extent on interviewer motivation and typing skills. Recent technological developments mean it is now possible for interviewers to digitally record respondents rather than typing as they speak, offering potentially substantial gains in data quality. In a recent NCRM study, we weighed the costs and benefits of implementing audio-recording of verbatim responses within a CAPI interview[1].

We used the 2012 Wellcome Trust Monitor[2], which is a random face-to-face survey on public attitudes to biomedical science. Respondents were randomly assigned to one of two experimental groups; in one group, verbatim answers to questions about scientific terms were audio-recorded, while in the other group they were typed by the interviewer in the conventional manner. Because it was necessary for ethical reasons to ask respondents for their consent to be audio-recorded, the design produced three groups: (1) those who gave consent and were allocated to the 'audio' group, so had their responses audio-recorded; (2) those who gave consent but were allocated to the 'typed' condition, so had their responses typed; and (3) those who did not give consent and regardless of whether they were allocated to the 'audio' or 'typed' condition, had their responses typed.

A key limitation of audio-recording is that a substantial minority, a third of respondents in this case, do not give consent to be recorded. Moreover, we found that the probability of giving consent was related to respondent characteristics. Those with higher educational qualifications, higher scientific knowledge, and those with a greater willingness to participate in research were the most likely to give consent. Thus, analysis of the raw response data would over-represent these groups relative to the general population.

Three indicators were used to compare the data quality of the responses between groups: the total number of words recorded, the number of 'thematic' codes applied by coders, and the average word length. The audio-recorded responses were significantly and substantially longer than the two typed groups. However, there were no significant differences with regard to the number of codes applied or the average word length.

We also examined whether audio-recording responses reduced between interviewer variability on these indicators, as audio-recording should, in principle, reduce the interviewer's influence on these outcomes. We found significantly lower between interviewer variance in both the number of words and the number of codes applied for the audio-recorded group compared to the typed groups, though not for the average word length.

In conclusion, audio-recording of verbatim responses appears to offer significant gains in data quality compared to requiring interviewers to type responses as they are articulated by respondents. However, this must be weighed against the substantial and non-random rate of refusal to be audio-recorded. Future research could usefully focus on methods for increasing the rate of consent to be recorded.

References

1 Sturgis, P and Luff, R (2015) Audio-recording of Open-ended Survey Questions: A Solution to the Problem of Interviewer Transcription? In: Engel, U (Ed) Survey Measurements: Techniques, Data Quality and Sources of Error. Frankfurt: Campus, pp.42-57.

2 Wellcome Trust Monitor: http://www.wellcome.ac.uk/About-us/Publications/Reports/Public-engagement/WTX058859.htm

How consistent is respondent behaviour to allow linkage to health administrative data over time?

Tarek Mostafa and Richard D Wiggins, University College London

Household surveys are increasingly being linked to administrative records with the potential of greatly enriching survey content on subjects such as health, education and income. One major challenge to data linkage is non-consent. Non-consent occurs when respondents refuse permission to link their administrative records to their survey data. This problem inevitably leads to information loss and possibly bias if consent is correlated with key characteristics of the respondents.

Despite the recent developments in the analysis of consent[2,3,4,5] the evidence is still scarce. The existing literature focused on the patterns of consent arising in cross-sectional surveys and very little is known about consent behaviour over time. Our study is one of the first longitudinal explorations of consent in the context of attempts to link survey and administrative data. It relies on a theoretical framework which distinguishes between passive, active, consistent and inconsistent consent behaviour and uses a series of models including multivariate probit analyses in order to identify the nature of consent as a latent trait and linear probability models which include interviewers as fixed effects.

The study exploits three consent questions to link health records held by the National Health Service (NHS) to the Millennium Cohort Study (MCS). The questions correspond to waves 1 (age 9 months), 2 (age 3 years) and 4 (age 7 years) of MCS. The questions are: To what extent is consent behaviour consistent over time? and Can consent behaviour be described as active or passive?

In order to answer the questions Cialdini's[1] framework on compliance in survey research is developed by testing four scenarios which sub-divide consistency/inconsistency in consent behaviour along the lines of activeness/passiveness. *Active consistency* is the case where respondents are aware of their previous choices and are committed to make the same choices on future occasions because of stable beliefs or personality traits (e.g. belief in the importance of scientific research, being a private person, etc). *Passive consistency* is the case where respondents make consistent choices over time even though the decision making process is passive. This means that consent decisions do not reflect an active adherence to well-defined beliefs but rather external influences such as the respondents' circumstances at the time of the interview and the impact of the interviewers. *Active inconsistency* is the case where respondents are aware of their previous choices and intentionally behave in inconsistent ways. This change in behaviour could be the result of a change in convictions. For instance, a past consenter might actively decide to withhold consent after a breach to data confidentiality. *Passive inconsistency* is the case where respondents switch from consenters to non-consenters or vice versa. This switch is not the result of changes in convictions but rather the result of changes to the circumstances of the respondent (e.g. divorce, acute health problems), changes to the interviewers over time (e.g. persistence in pursuing consent), and the fact that respondents could have forgotten what they did in the past. In all cases, the respondent has a passive role and the changes in consent behaviour are caused by extrinsic factors.

In summary the analyses provide evidence in support of passive consistent behaviour. Firstly, 77% of respondents behaved consistently over time by either consenting or refusing to consent in all waves. Secondly, the cross-equation correlations from the multivariate probit models showed that the unobserved parts of the consent outcomes are weakly associated over time, and therefore, cannot really be held to indicate the existence of strongly held latent convictions about consent. Thirdly, the likelihood of consent and the likelihood of switching behaviour over time are related to the respondents' circumstances, and to the variation the impact interviewers have on the respondents' willingness to consent. These three findings indicate that, for the majority of respondents, consent is not driven by personal convictions but rather depends on the circumstances of the respondent at the time of the interview and on the potential influence of the interviewers.

In sum when it comes to using linked survey and administrative data, users need to take into account the potential sample bias resulting from the correlations between the respondents' characteristics and the agreement to consent. On a practical level the longitudinal evidence based on our project suggest that it is important to brief interviewers about what to expect and encourage them to adapt to the respondent's circumstances when attempting to gain their cooperation to consent to record linkage.

References

1 Cialdini, R. Wosinska, W. Barrett, D. Butner, J. and Gornik-Durose, M. 1999. Compliance with a Request in Two Cultures: The Differential Influence of Social Proof and Commitment/Consistency on Collectivists and Individualists. Personality and Social Psychology Bulletin 25, 1242-1253.

2 Korbmacher, J. and Schroeder, M. 2013. Consent When Linking Survey Data with Administrative Records: The Role of the Interviewer. Survey Research Methods 7, 115–131.

3 Mostafa, T. 2015. Variation within Households in Consent to Link Survey Data to Administrative Records: Evidence from the UK Millennium Cohort Study. International Journal of Social Research Methodology, published online on: 23 March 2015.

4 Sakshaug, J. Couper, M. Ofstedal, M. and Weir, D. 2012. Linking survey and administrative records mechanisms of consent. Sociological Methods and Research 41, 535-569.

5 Sala, E. Burton, J. and Knies, G. 2012. Correlates of obtaining informed consent to data linkage: respondent, interview and interviewer characteristics. Sociological Methods and Research 41, 414–439.

Latest research methods audio and video

NCRM video and audio podcast series features a wide range of research methods related topics discussed by experts from NCRM affiliated projects and events.

Latest audio podcasts:

Care data: the challenges of linking health service data; Neil Serougi; 19 July 2015

Data linkage: challenges and opportunities; Peter Elias; 19 July 2015

Teaching and learning social research methods; Melanie Nind, Daniel Kilburn and Rebekah Luff; 26 May 2015

Using Skype in qualitative interviews with young people; by Susie Weller; 28 April 2015

The audio podcast is available on www.ncrm.ac.uk/resources/podcasts

Latest video:

Data linkage and the ESRC; Peter Elias; 19 July 2015

Civil society and data linkage in the age of big data; Rob Procter; 19 July 2015

Exploring the Social and Ethical Issues in Data Linkage; Mandy Chessell; 19 July 2015

Civil liberties and data linkage; Ian Dung; 19 July 2015

The video is available on www.ncrm.ac.uk/resources/video or on the NCRM National Centre for Research Methods YouTube channel

ABOUT NCRM

The ESRC National Centre for Research Methods (NCRM) is a network of research groups, each conducting research and training in an area of social science research methods.

NCRM brings together researchers from across the UK with a wide range of research methods expertise, at the frontiers of developments in research methodology.

NCRM disseminates innovations and developments in research methods through training courses and events and through other direct engagement with researchers, but also by cooperating with other organisations and initiatives with an interest in social science research methods.

NCRM was established in 2004 as part of the Economic and Social Research Council's (ESRC) strategy to improve the standards of research methods across the UK social science community. NCRM acts as a strategic focal point for developments in research, training and capacity building related to research methods, both at the national level and cutting across social science disciplines.

For more information about the NCRM and its activities please see our website http://www.ncrm.ac.uk

MethodsNews is published three times a year by the National Centre for Research Methods.
Editorial team: Melanie Nind and Eva Nedbalova

National Centre for Research Methods
Social Sciences
University of Southampton
Southampton SO17 1BJ
United Kingdom

Email info@ncrm.ac.uk
Tel +44 23 8059 8199
Web http://www.ncrm.ac.uk
Twitter @NCRMUK

NCRM National Centre for Research Methods

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL